

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

04

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

RECEIVED BY
NASA STI FACILITY

DATE:

DCAF NO. 103295

PROCESSED BY

☒ NASA STI FACILITY

☐ ESA - SDS ☐ AIAA

76-10504

CR-148824

REMOTE SENSING IN PERU BY
MEANS OF LANDSAT IMAGERY

(E76-10504) DIGITAL PROCESSING OF SATELLITE
IMAGERY APPLICATION TO JUNGLE AREAS OF PERU
Final Report, 1973 - 1975 (Instituto
Geofisico del Peru) 84 p HC \$5.00 CSCL 05B

N76-32616

G3/43 00504
Unclas

PREPARED FOR:

NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

SPONSORED BY:

INSTITUTO GEOFISICO DEL PERU

OFICINA NACIONAL DE EVALUACION
DE RECURSOS NATURALES

UNIVERSIDAD NACIONAL AGRARIA

1302A

RECEIVED

SEP 08 1976

SIS/902.6

REMOTE SENSING IN PERU BY
MEANS OF LANDSAT IMAGERY

Original photography may be purchased from:
EROS Data Center
10th and Dakota Avenue
Sioux Falls, SD 57198

Prepared for:

National Aeronautics and
Space Administration

Sponsored by:

Instituto Geofísico del Peru

Oficina Nacional de Evaluación
de Recursos Naturales

Universidad Nacional Agraria

**ORIGINAL CONTAINS
COLOR ILLUSTRATIONS**

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. ARSI - 76		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle DIGITAL PROCESSING OF SATELLITE IMAGERY APPLICATION TO JUNGLE AREAS OF PERU"				5. Report Date Feb. 24, 1976	
				6. Performing Organization Code	
7. Author(s) José C. Pomalaza -- ARSI Carlos A. Pomalaza - Jorge Espinoza				8. Performing Organization Report No. IGP-ARSI - 76	
9. Performing Organization Name and Address INSTITUTO GEOFISICO DEL PERU Apartado 3747 Ave. Armendariz 497 Miraflores - Lima - Peru				10. Work Unit No.	
				11. Contract or Grant No. SR - 302	
12. Sponsoring Agency Name and Address Instituto Geofísico del Perú Oficina Nacional de Evaluación de Recursos Naturales Universidad Nacional Agraria				13. Type of Report and Period Covered Final Report 1973 - 1975	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract Computerized techniques are developed to process CCTs from LANDSAT-1 Satellite. A description of the set of program, named INTERPRET, is made and its application to a classification work in areas of the Peruvian Jungle is presented.					
17. Key Words (Selected by Author(s)) Digital Processing Application Pattern Recognition Natural Resources Automatic Classification				18. Distribution Statement	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages	
				22. Price*	

*For sale by the Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia 22151.

Contributors:

José C. Pomalaza	Principal Investigator (IGP)
Medardo Molina	Coinvestigator (UNA)
Pedro Lavi	Coinvestigator (ONERN)
Eduardo Seminario	UNA
Walter Danjoy	ONERN
Jorge Espinoza	IGP
Arturo Pomalaza	IGP

TABLE OF CONTENTS

FOREWORD. José C. Pomalaza.

"DIGITAL PROCESSING OF SATELLITE IMAGERY, APPLICATION TO JUNGLE AREAS OF PERU". José C. Pomalaza, Carlos A. Pomalaza, Jorge D. Espinoza.

"A FORESTRY STUDY IN THE CONTAMANA-TARAPOTO AREA BY MEANS OF LANDSAT I IMAGERY". Walter Danjoy.

Foreword

When the ERTS project was officially announced by NASA it was received with great interest in Peru since it represented a way to develop Remote Sensing Techniques, that could substantially accelerate and improve the study of our natural resources, without the higher initial investment necessary to acquire airborne multispectral sensors.

The Geophysical Institute of Peru (IGP) and the National Office for Evaluation of Natural Resources (ONERN) joined their efforts for Remote Sensing development and with the help of the National Agrarian University prepared a proposal to participate in the ERTS project and presented it to NASA in 1971.

Since then an intense program of activities was developed and in spite of initial difficulties it has evolved toward a successful program.

Previous to the launching of the first ERTS satellite, technical groups were organized to work on the ERTS imagery that would be arriving, all the basic information that was available on the Santa River Basin was collected and at the IGP, Mr. Eduardo Seminario under the direction of Dr. Medardo Molina from UNA made a Hydrologic Study of the Santa River Basin to support the studies with Landsat Imagery.

In order to promote the national interest on remote sensing and increase support to this project, talks were given at Universities ,

national meetings and government organizations. These efforts culminated with a Seminar on Remote Sensing given at UNA in April 1972.

In July 1972, an agreement between ONERN and IGP was signed to work the development of Remote Sensing in Peru and to seek national or foreign financial aid to support it.

ERTS-1 was successfully launched in August 1972, after a number of postponements, and finally the first images started to arrive during the first half of 1973.

The images were taken over our observation site on the Amazon Jungle. At that time oil exploration activities in the jungle were booming and a priority interest was given by the government to that region of the country. It was then decided that the efforts of the teams at IGP, ONERN and UNA will concentrate their work on studies about the jungle areas covered by ERTS-1.

Mr. Eduardo Seminario and Dr. Molina made a study of Meanders over rivers in the Amazon Jungle. By the end of 1973 Mr. Seminario was leaving to Europe to benefit from a scholarship and Mr. Molina took a leave of absence in early 1974 to work for UNESCO.

At IGP an infrastructure was implemented to deal with the imagery coming from NASA, to enlarge the black and white 70 mm. and to produce 35 mm color composites.

The imagery received from NASA was given broad diffusion being the most significant applications the use of 250,000 enlargements as a help for a photomosaic work at Servicio Nacional de Aerofotografía (SAN), the use of 1:250,000 enlargements by Belco Petroleum Co. in their planning of an oil pipeline in the Northern peruvian jungle. The use of 1:500,000 enlargements by Minero Peru in their research work.

In 1974 under the auspices of the National Institute of planning a proposal by ONERN and IGP was presented to the Canadian International Development Agency (CIDA) to finance a pilot Remote Sensing project. The preliminary part of this project is already underway and the main part is expected to start this March with an study based on Landsat 1 and 2 imagery.

This report presents the work that was performed at IGP, ONERN and UNA during the period of 1972 through 1975 with the Landsat 1 (ERTS-1) imagery.

During the course of this work it was decided that ONERN will concentrate their efforts in the development of photointerpretation techniques for multispectral imagery as a natural evolution of their work on the evaluation of natural resources of the country. IGP will work mainly in computerized methods of data handling and analysis. UNA concentrated their efforts in hydrologic studies.

"DIGITAL PROCESSING OF SATELLITE IMAGERY
APPLICATION TO JUNGLE AREAS OF PERU"

By

José C. Pomalaza
Carlos A. Pomalaza
Jorge D. Espinoza

1976

Geophysical Institute of Perú
Ancón Observatory

A B S T R A C T

Computerized techniques are developed to process CCTs from Landsat-1 Satellite. A description of the set of program, named INTERPRET, is made and its application to a classification work in areas of the Peruvian Jungle is presented.

I N D E X

	<u>Page</u>
Illustrations	iii-2
Tables	v-2
I. Introduction	2-1
II. Report Summary	2-3
III. The Problem of Machine Classification of Remotely Sensed Data	2-3
IV. The Clustering Method	2-6
V. The Computer Compatible Tape (CCT)	2-7
VI. The Development of a Computer Technique to Analyze Remotely Sensed Data Recorded on CCT's	2-8
VII. Applications of INTERPRET to the Study of the Amazonian Jungle	2-28
VIII. Conclusions	2-39
IX. Future Work	2-42
Acknowledgement	2-46

I L L U S T R A T I O N S

<u>Title</u>	<u>Page</u>
1. - Classification of Remotely Sensed Data	2-4
2. - Typical MSS Scene from LANDSAT-1 Channel 5	2-9
3. - Computer System Used	2-11
4. - Program ALMAP I - Example of messages interchanged between the computer and the user	2-13
5. - Map of Ucayali River. Program ALMAP I, character re- fer to different gray levels in channel 6. Vertical numbers indicate scan lines	2-14
6. - Map of Ucayali River West of Contamana. Program ALMAP II.	2-15
7. - Flow Chart of ISODATA	2-17
8. - Typical Result from CLUST. Subclusters are given in a table indicating their most important parameters	2-20
9. - Typical Messages Between the User and the Computer. Program CLASIF	2-22
10. - Flow Chart of the Process to Determine Sample Data Points from CCTs	2-24
11. - Flow Chart for Subcluster Formation and Optimization	2-25
12. - Flow Chart for Computer Classification. Continua- tion from Previous Picture.....	2-26
13. - Scene from LANDSAT-1 Chose for this Study. Zone I and Zone II are the Area Covered	2-29
14. - Map from an Early Version of INTERPRET. Multiple print was used to obtain a gray level presentation for different classes. The area covered is over the swamps inside Zone I of Fig. 13.	2-31

Illustrations (Cont.)

	<u>Page</u>
15. - Color Map from an Early Version of INTERPRET. The following color code was used:	2-33
BLACK	Highly Sedimented Water
BLUE and VIOLET	Water with Intermediate Degree of Sedimentation
ORANGE	Herbaceous Vegetation
GREEN	Mixture of Water and Vegetation or Wet Areas.
DARK YELLOW	Swamps
16. - Oscilloscope Maps from Present Version of INTERPRET	2-34
17. - Teleprinter Map from Present Version of INTERPRET	2-35
Special character were used for a gray level presentation	
18. - Color Map from Present version of INTERPRET The following colors were used:	2-36
YELLOW	Alluvial Forest Type III
GREEN	Mixture of Vegetation and water Wet Areas or Alluvial Forest - Type II
BLUE	Water
PINK	Swamps
WHITE	Unrecognized Patterns
19. - Extract from "Mapa Forestal del Perú 1976". Scale 1:1,000,000 National Agrarian University (UNA) ..	2-38
20. - Preliminary Gray Level Map of Tarapoto..... Brighter points represent urban areas. Medium level points represent unpaved roads and weak level points represent water.	2-40
21. - Future Computer System for Machine Analysis of Imagery	2-43

T A B L E S

	<u>Page</u>
TABLE I : ISODATA Main Functions as specified in Fig. 7	2-18/2-19
TABLE II : SYMBOLS Used on Flow Charts	2-23

I. Introduction

In the last eight years a great deal of effort have been spent in oil exploration activities in the peruvian amazonic jungle. At this time a pipeline that will bring the oil from the northern fields of the jungle to the coast is under construction. With the effects of the oil crisis still present, is not difficult to understand the interest of Peru in this vast region of the country.

With the urgency to develop our capability to produce oil came the necessity to obtain basic information about the jungle. The first images from ERTS-1 on that area showed that a great deal of important information could be obtained from them. The Geophysical Institute of Peru (IGP), the National Office for Evaluation of Natural Resources (ONERN) and the Agraria University (UNA) started working with those ERTS-1 images, ONERN and IGP worked to develop classification methods for multispectral imagery and UNA in the study of meanders in the Amazon Basin.

The first task was to select imagery with good quality, low cloud cover and with a large variety of interesting features. The image selected was the one taken over the region close to the cities of Contamana and Tarapoto.

To develop classification schemes, photographs from airplanes were used along with enlargements to 1:250,000 scale and color composites of 1:1,000,000 scale. In the course of this work it was found that significant detail was being lost in the photo interpretation process. The problem was mainly due to

one of the following causes: the photo interpreter found difficulties to switch from one channel to other to identify a feature, the details were too small on the 1:1,000,000 color composites or too faint in the enlargements, the photo interpreter got tired and started making errors.

It was then decided that computerized techniques should be studied as a way to solve some of these problems and to accelerate the interpretation process. The initial results at IGP were promising and then a more systematic effort was started in this field. It was decided that ONERN will continue working with photo interpretation methods and IGP will develop computer interpretation techniques, both works were carried out in parallel.

The purpose of this report is to present what was accomplished at the IGP to develop computer techniques to process computer compatible tapes (CCT) from Landsat Satellites.

This work concentrated in the multispectral analysis of CCTs, the problem of geometric correction and application of techniques to define geometric features and temporal variations was left for future development.

When this work started a Datacraft 6024 computer was available at IGP, the memory capacity was 32 K bytes and the teleprinter was too slow for practical use which made the progress too difficult.

Then, it was decided to use the IBM 360/40 computer at

the National University of Engineering, but it was found that a large computer center is not particularly suitable for the type of work that we were interested in. Finally a Hewlett Packard 2100 computer system installed by Stanford Research Institute at the Ancon Observatory of IGP became available and then it was possible to satisfy most of the requirements for our work and the programs for machine processing of CCTs reached a more satisfying status.

II. Report Summary

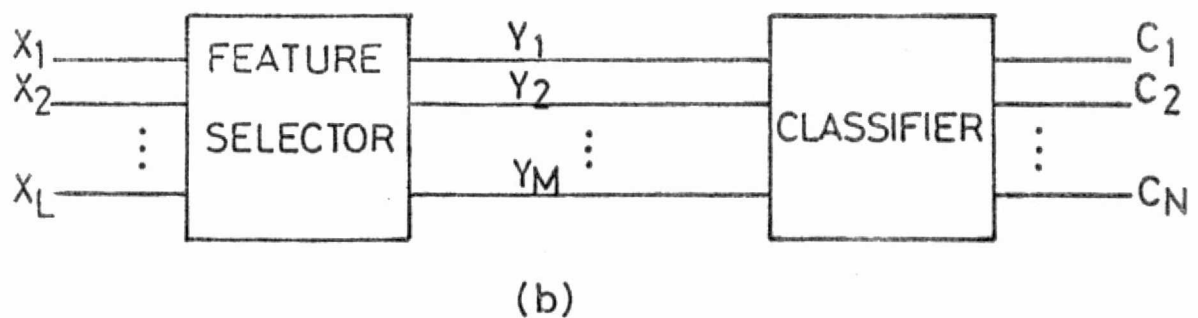
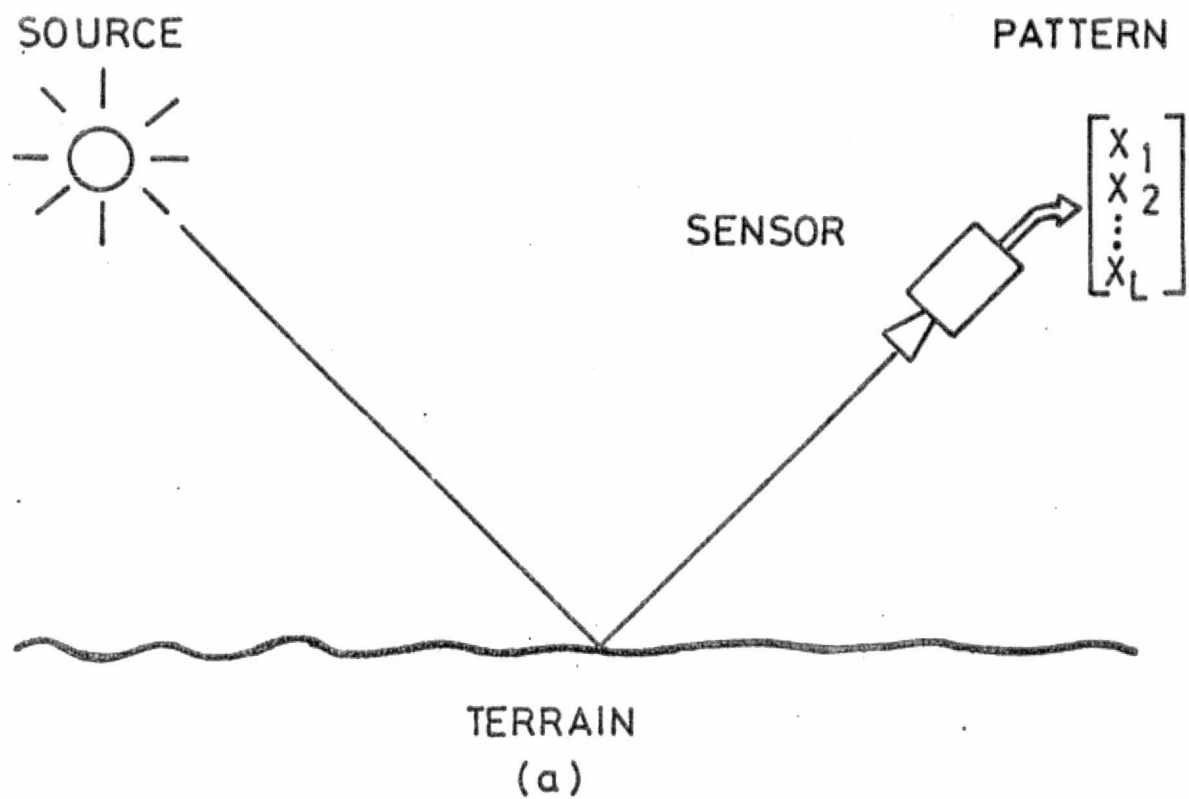
In chapters III and IV the general problem of machine classification is introduced followed by a brief description of CCTs in chapter V.

The computer techniques that were developed during this work are presented in chapter VI and their application to the study of jungle areas are described in chapter VII.

The conclusions that were reached are presented on chapter VIII, and finally chapter IX describes our plans for future work.

III. The Problem of Machine Classification of Remotely Sensed Data

Fig. 1 illustrates the general problem of machine classification. The source illuminates the terrain, the sensor receives the reflected radiation and at a particular time generates a set of numbers, patterns, which are a measurement of the received ra-



(b)
FIGURE 1

CLASSIFICATION OF REMOTELY SENSED DATA
INTO N CLASSES

diation.

The feature selector extracts those numbers that are significant to define or characterize the surface under observation. This new set can be represented by a vector generally known as the feature vector.

The purpose of the classifier is to assign each input pattern to one of the possible pattern classes. The classifier makes this assignment on the basis of the feature vector obtained from the input patterns and also based on past experience.

The development of a classifier has been the main objective of this work and the feature extractor has been only considered as a particular case of the classifier.

The problem of classification is basically one of partitioning of the space defined by the feature vectors in regions, one region for each class. The partition is performed in such a way that the probability of missclassification is minimized. Extensive literature exists on the various methods to solve this problem^{4,5}.

At present there are a number of methods for machine classification^{4,6}, they are based on the use of the covariance matrix, in the measurement of probability distributions (parametric and non parametric) clustering techniques and other. The method chosen will depend in general on the problem itself and on the avail-

*Superscripts indicate references at the end of report.

lable digital equipment.

The main purpose of this work was to develop computer techniques to analyse Landsat-1 imagery over the peruvian Amazonic Jungle close to Ucayali River, one of the Amazon tributaries. The structure of the multiespectral data obtained was very complex, obviously non gaussian, ground truth is almost nonexistent and difficult to obtain because of the inaccessibility of the region. The digital equipment used in the problem was based on a Hewlett Packard 2100 computer with 32 k bytes of memory, two tape units and a 600 character/min teleprinter. These facts led us to choose clustering techniques as the base for our classification algorithms.

IV. The Clustering Method

Clustering techniques are based on the determination of Natural groupings in a set of data. The space defined by the feature vectors is structured and there are regions where the density of points is greater than the surrounding space, clustering methods permit to define volumes that will enclose those regions.

Intuitively, one can assume that data points, or feature vectors, corresponding to objects on the terrain that are similar should come close in the feature space and hopefully forming clusters.

It is fundamental to the clustering method the definition of a metric to determine closeness, or similarity, in the feature space. From the variety of similarity measures⁴, the Euclidean -

distance between feature vectors was used in this work and closeness was defined in base of the nearest neighbor¹¹ criterion.

V. The Computer Compatible Tape (CCT)

As a new tool for the study of remotely sensed land surfaces come the CCTs¹ generated from data collected by the multispectral scanners (MSS) on board of Landsat Satellites.

The MSS covers the visible and near infrared spectrum with four channels:

Channel

4	0.5 - 0.6 microns
5	0.6 - 0.7 "
6	0.7 - 0.8 "
7	0.8 - 1.1 "

The MSS scans cross track swaths 185 km wide at a nominal altitude of 494 nautical miles. It consists of six detectors for each of the four channels, video outputs from each detector are sampled, commutated and multiplexed. The commutated samples are encoded and transmitted to ground-base receiving sites. The receiving stations compile the raw data on video tapes and transmit these tapes to the NASA Data Processing Facility (NDPF) at the Goddard Space Flight Center, Greenbelt Maryland. The NDPF corrects, calibrates and formats the raw MSS data and converts it to binary form on Computer Compatible Tapes (CCT).

A 185 km square ground scene on the CCT is a final pro-

duct of the MSS. Fig. 2 illustrates a typical scene corresponding to a pass of Landsat-1 Satellite over the Ucayali and Huallaga River in the Central Amazon Jungle of Peru.

A scene is made up of parallel scan lines each containing a large number of video points or picture elements (Pixels).

The NDPI records a complete ground scene on four separate CCTs. Each CCT contains image data for one 46.25-185 Km strip. The CCT contains data that has been radiometrically calibrated¹ but not geometrically corrected for skew due to earth rotation. CCTs come in 800 bpi and 1600 bpi densities of tapes.

In this work 1600 bpi CCT corresponding to the image shown in Figure 2 was used.

VI. The Development of a Computer Techniques to Analyse Remotely Sensed Data Recorded on CCTs

As a framework for this work the following basic requirements were defined:

- ① Clustering techniques will be used.
- ② The programs will be interactive so that they permit, as much as possible within the limitations of our equipment, the intervention of the specialist during the analysis.
- ③ The programs should use the four channels from the CCTs but will have, as an option, the use of less num

2-9



FIG. 2 TYPICAL MSS SCENE FROM LANDSAT I
CHANNEL 5

ber of channels.

- The presentation of results will be, either on teleprinter or CRT.
- Although they will be developed with a particular problem in mind, they should be sufficiently general so that they can be used in other applications too.
- The programs should use, as much as possible, standard languages, like ALGOL or FORTRAN IV.

Fig. 3 illustrates the equipment used in this work.

The group of programs developed to process CCTs was named INTERPRET and consist of the following:

- Subroutines for data manipulations from CCTs, written in ASSEMBLER language callable from ALGOL and FORTRAN IV programs.
- Programs ALMAP I and ALMAP II display on the teleprinter and oscilloscope respectively the video information, levels of reflectance, from one of the channels stored on CCTs. The display is in the form of a selected area from a scene. The language used is ALGOL.

These programs have an option that permits to limit the range of the video information so that, only those areas on the terrain that produce reflexions levels within given limits will be displayed.

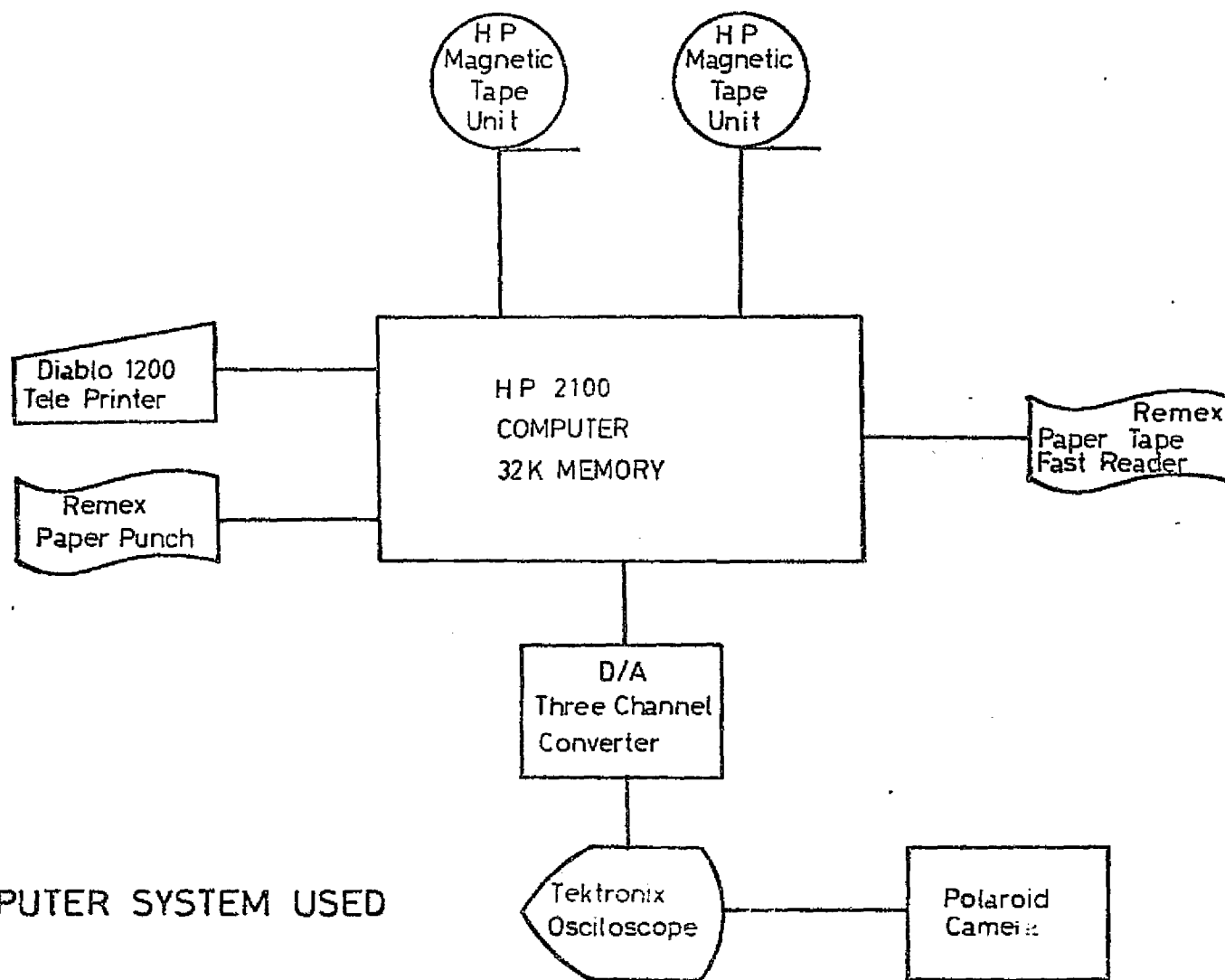


FIG.3- COMPUTER SYSTEM USED

played. Also the desired range, between the limits can be divided into slices of equal width and, displayed with different symbols on the teleprinter or different gray levels on the Oscilloscope, this permits to identify features on the terrain with levels on the video information.

It was found desirable to be able to vary the scale on the display, and therefore ALMAP I and II were provided with an option that permits to produce displays at a prescribed scale.

The programs are interactive through a continuous interchange of messages between the machine and the operator. Fig.4, shows a sample of ALMAP I messages. Fig.5 and 6 show typical displays from ALMAP I and II respectively.

Programs HISTA and HISTB display, on the oscilloscope and teleprinter respectively, two dimensional histograms of any pair of MSS channels. The value of histogram elements are indicated by means of gray levels on the oscilloscope or symbols on the teleprinter. These programs are interactive and have options that permit the user to display a prescribed range of levels or to slice out the more intense peaks so that to enhance the fine structure of the histograms.

```

ALGOL MODIFIED MAP PROGRAM (ANCON SEPTEMBER 75)
UNIT,TAPE?,RECORD LENGHT?,RECORD POSITION?
9.3296,-1
TITLE?
NORTE DE CONTAMANA
CHANNEL?
6
INITIAL LINE?,FINAL LINE?,INITIAL POINT?,FINAL POINT?
250,320.200,260
LINE INTERVAL?,POINT INTERVAL?
1.1
MINIMUM LEVEL?,MAXIMUM LEVEL?
0.33
LEVEL INTERVAL?
3
CHARACTERS?
SI
00000000111
123456789012
abcdefghijkl
NORTE DE CONTAMANA
CHANNEL: 6
FROM TO CHAR
0 2 a
3 5 b
6 8 c
9 11 d
12 14 e
15 17 f
18 20 g
21 23 h
24 26 i
27 29 j
30 32 k
33 33 l

```

FIG. 4 - Program ALMAP I. Example of messages
interchanged between the user and the
computer.

6-11-57

ORIGINAL PAGE IS
OF POOR QUALITY



FIG.6 - MAP OF UCAYALI RIVER WEST OF CONTAMANA

PROGRAM "ALMAP II"

ORIGINAL PAGE IS
OF POOR QUALITY

- Program CLUST is the main one, it permits to define clusters in the feature space and has been developed based on the technique named ISODATA (Iterative Self Organizing Data Analysis), Reference 5.

ISODATA is a collection of iterative techniques that permit to find the principal modes, or subclusters, in the data by grouping those patterns that are close together, and representing them by a hypersphere centered on the average of the patterns.

By means of these techniques a cluster is represented by a number of subclusters in such a way that the variance of each subcluster is kept smaller than a parameter defined by the user. In each iteration the subclusters are split or lump together depending on a set of control parameters. Fig 7 and Table I briefly describe ISODATA.

CLUST has been implemented in ALGOL with a modular structure, it consists of a set of subroutines that perform each of the functions described in Fig. 7. CLUST is interactive and permits the operator to redefine all the control parameters after a given number of iterations.

The results obtained by CLUST are a set of subclusters which are printed in form of a table specify-

2-17

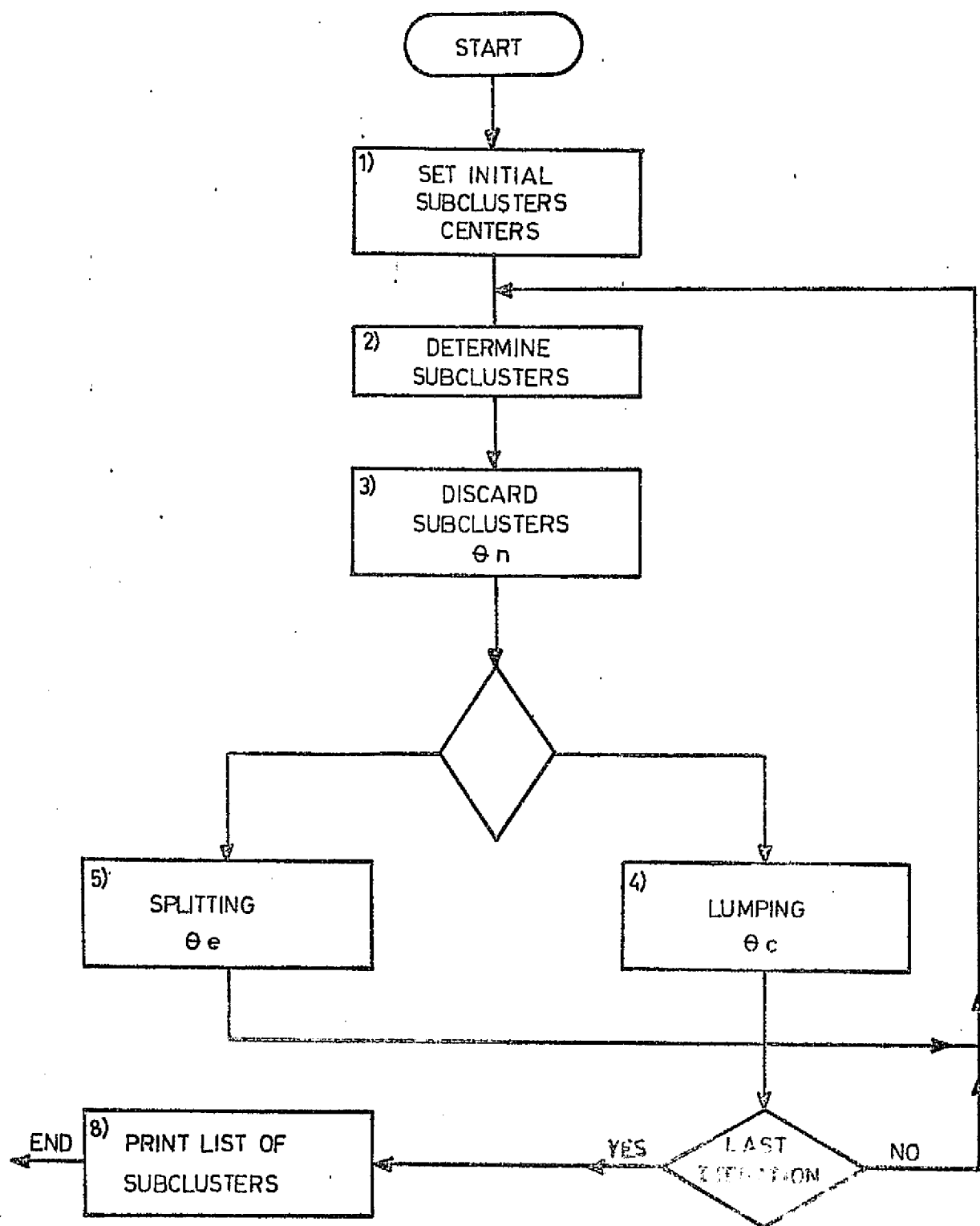


FIG. 7 - FLOW CHART OF ISODATA SEE TABLE N. 1

TABLE IISODATA Main Functions as Specified in Fig.7

Step	Function Description
1	Initial subcluster centers are fed to the computer and the control process parameters are specified.
2	<p>Input patterns are allocated to the subclusters using a minimum distance criterion.</p> <p>The following variables are computed:</p> <ul style="list-style-type: none"> - Average vector for each subcluster, - Standard deviation for each sub--cluster. - Average distance of the patterns to their subcluster center - Number of patterns for each subcluster.
3	Discard those subclusters whose number of patterns is less than the parameter Θ_n .
4	<p>Take this way if:</p> <ul style="list-style-type: none"> - Last iteration - Or even iteration - Or number of subclusters greater than two times the number of subclusters(2 N)

T A B L E I

(Continuation)

Step	Function Description
5	<p>specified at the beginning of the process.</p> <p>Take this way if:</p> <ul style="list-style-type: none">- Odd iteration- Or number subclusters less than half the number of subclusters ($0.5 N$) specified at the beginning of the process.
6	<p>Lump together two subclusters if the distance between their average vectors is less than θ_c.</p>
7	<p>Split a subcluster in two if in any dimension the standard deviation is greater than θ_e.</p>

TARAPOTO 5 DICIEMBRE 1975
 ZONAS DE MUESTRA?
 SI
 UNIT NUMBER?(TAPE READER=5, TELEPRINTER=1)
 3
 NUMERO DE ZONAS DE MUESTRA?
 1
 ZONAS DE MUESTRA:
 2,X1,X2,X3,XX
 X = LINEA DE LA ZONA (X < 0 SIN DE GATOS)
 X1 = LINEA INICIAL
 X2 = LINEA FINAL
 X3 = PUNTO INICIAL
 X4 = PUNTO FINAL
 1,2749,2762,139,155
 -1
 LISTADO?
 SI

# ZONA	LINEA INICIAL	LINEA FINAL	PUNTO INICIAL	PUNTO FINAL
1	2749	2762	139	155

 HAS DATA?
 NO
 UNIT NUMBER?(SCRATCH TAPE ONE)
 10
 HOW MANY CHANNELS? WHAT CHANNELS?
 2,5,7
 UNLOAD ERTS TAPE, LOAD SCRATCH TAPE TWO
 SUBCLUSTER CENTERS?
 SI
 INPUT UNIT?(PRINTER=1, TAPE READER=5)
 1
 NUMERO DE SUBCLUSTERS?
 2
 N,XX,XX,XX,XX,XX,XX,XX,XX
 N=NUMERO DEL SUBCLUSTER
 XX,XX COORDENADAS DEL SUBCLUSTER CENTER
 PARA ACABAR N<0
 1,32.1,14.1
 2,33.0,16.0
 -1
 LISTADO?
 SI

# SUBCLUSTER	X1	X2
1	32.10	19.10
2	33.00	16.00

 HAS DATA?
 NO
 PARAMETERS?
 SI
 TETAC?, TETA?, TETAN?, LP, MITER?, NDCLU?
 2,5,1,0,0,5,0,12
 TARAPOTO 5 DICIEMBRE 1975
 NOMENCLATURE:
 NI=POPULATION
 X=AVERAGE VECTOR COORDINATES
 AV=AVERAGE DISTANCE FROM THE AVERAGE VECTOR
 S=STANDARD DEVIATION

#	NI	X1	X2	S1	S2	AV
1	26	33.5385	24.1539	1.2770	1.0007	1.5310
2	2	51.5000	22.5000	2.5000	1.5000	2.0155
3	13	28.6923	27.0760	.0213	.0166	1.1676
4	23	35.3479	22.3912	1.0047	1.2056	1.4377
5	48	33.5417	20.9542	1.2061	.0005	1.4377
6	9	42.5556	21.0000	.0060	.6667	.7756
7	17	26.0000	26.0235	.7670	1.1497	1.2052
8	5	44.2000	23.0000	.0700	1.1662	1.4363
9	36	20.6944	21.1111	1.0755	1.0402	1.3033
10	13	25.3077	22.2309	1.1358	.7994	1.3105
11	22	20.1364	24.0909	.0103	.7925	1.0092
12	10	39.5000	18.0000	.5000	1.1662	1.1040
13	9	22.5556	27.0000	1.0657	1.1967	1.5153
14	9	43.3750	18.2500	1.1110	.8292	1.2340
15	7	45.4286	21.0000	.7204	.0250	.9395
16	17	36.6471	19.5000	1.0815	.9113	1.3436
17	5	22.0000	25.0000	.6325	.6325	.6020

 OVERALL AVERAGE DISTANCE OF PATTERNS: 1.32557
 TOTAL POPULATION: 270

FIG.8 - Typical result from CLUST. Subclusters are given in a table indicating their most important parameters.

ORIGINAL PAGE
 OF FOUR

ing their main characteristics such as feature space coordinates, population, standard deviation and radius. Fig. 8 shows a typical print out from CLUST.

- . Program CLASIF written in ALGOL performs the classification based on the subclusters defined by CLUST and selected through an optimization process that will be described later.

A pattern is assigned to a subcluster if its Euclidean distance to that subcluster is the minimum distance to any subcluster and it is smaller than a maximum radius defined during the optimization process. The result of this classification can be displayed by the teleprinter or the oscilloscope.

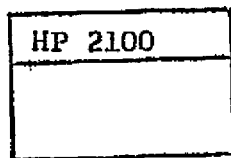
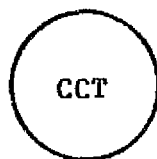
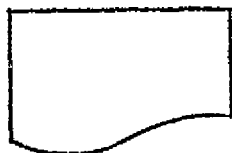
A modification was made to the teleprinter symbols so that to produce a gray level presentation on the teleprinter.

- . The program is interactive and permits to change the area to classify and the subclusters parameters. Fig. 9 shows a sample of CLASIF messages.

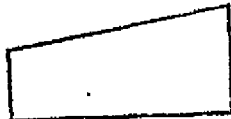
With these set of programs it was possible to develop a methodology to perform the computer analysis of CCTs. Figures 10, 11 and 12 describe the process of analysis in which the user plays a central role.

TABLE II

Symbols used on flow charts

HP 2100 Computer
32 K BytesMagnetic Tape unit
Loaded with a CCT

1200 Diablo Printer



1200 Diablo Keyboard

Tektronix Oscilloscope
Equipped with a Polaroid.
CameraRemex Fast Paper Tape Reader
and PunchSpecialist using
INTERPRETREPRO.
ORIGINAL

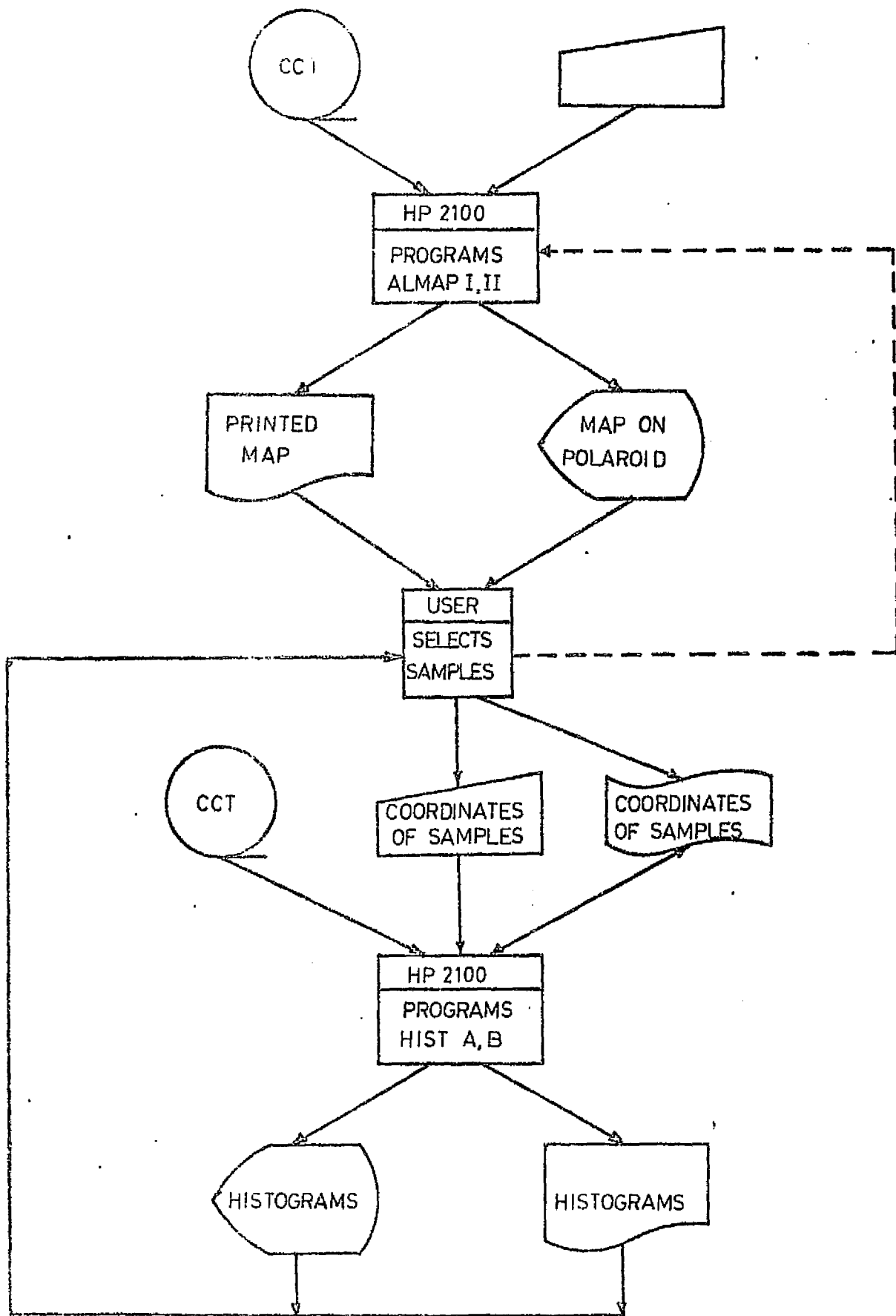


FIG.10- FLOW CHART OF THE PROCESS TO DETERMINE
SAMPLE DATA POINTS FROM CCTS

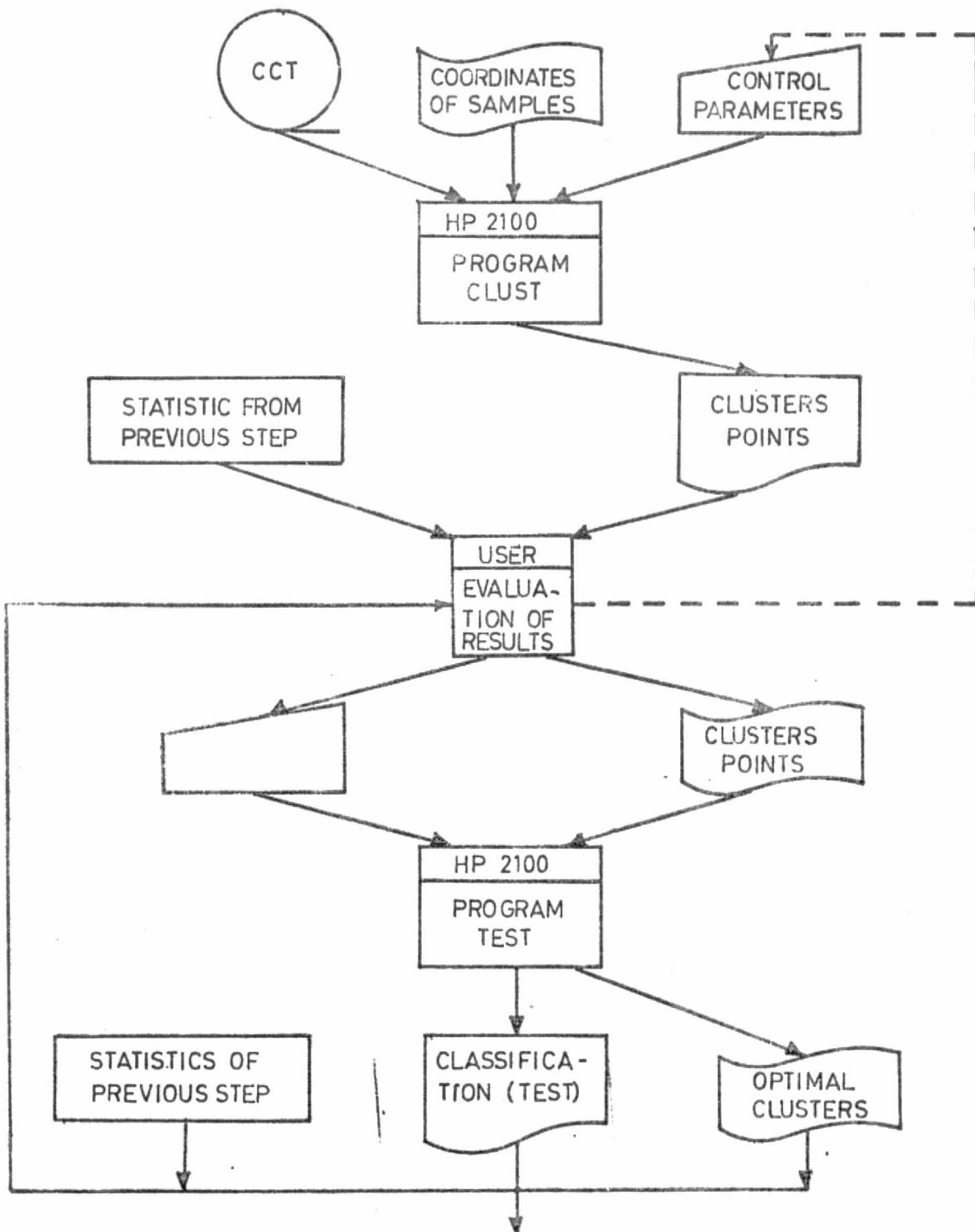


FIG. 11- FLOW CHART FOR SUBCLUSTER FORMATION AND OPTIMIZATION

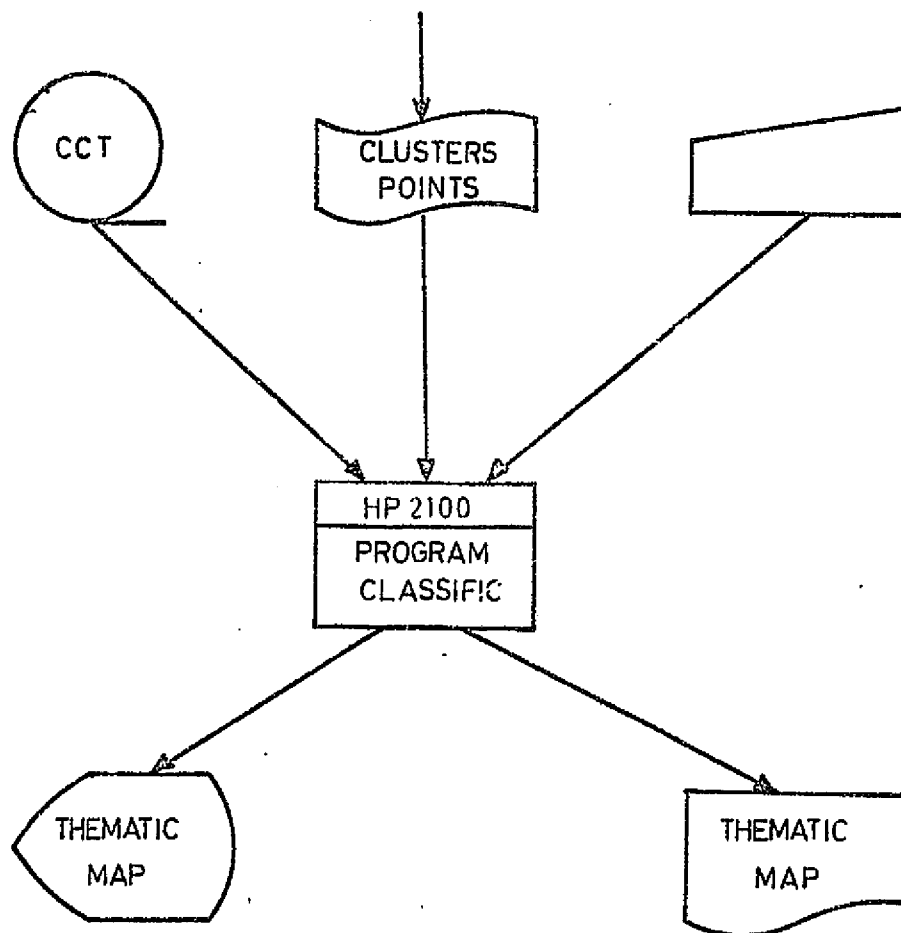


FIG 12 FLOW CHART FOR COMPUTER CLASSIFICATION
CONTINUATION FROM PREVIOUS PICTURE

Once the problem has been defined the process of analysis starts with the selection of samples. The user helped by all the information available, aerophotography and ground thruth, selects areas with known characteristics by displaying the Landsat information on the teleprinter by means of ALMAP I.

The user determines the coordinates of the samples and ask the computer to punch them on paper tape. With the help of HISTA or HISTB the user display on the teleprinter, or on a polaroid picture, two dimentional histograms obtained for the sample areas, later these histograms will be used to check the results of CLUST.

The second step of the process is to determine subclusters from the sample data by means of CLUST. Once the subclusters are determined they are mapped on two dimentional histograms and compared with the histograms determined in the previous step.

The sample patterns that are fed to CLUST are not completely "pure", that is, they contain data that do not belong to a known classification and therefore those patterns will be rejected during this step.

Next the subclusters are passed through an optimization process by means of the program CLASIF. Each subcluster is tested using the classification algorithm of CLASIF and the resulting maps are displayed on the oscilloscope, and on the teleprinter. Data from a known area, different from that of the sample data, is used for this process. The radius of each subcluster is reduced to minimize the number of missclassifications or increase to maximize

correct classifications.

The last step is the analysis of the area of interest, the classification is performed using CLASIF and the resulting maps are either displayed on the teleprinter or on oscilloscope as illustrated on figure 12. If the maps are not satisfactory then new sampling areas are defined and the process is repeated again but only limited to those classes that were not being correctly classified. The process usually does not require more than two iterations.

VII. Applications of INTERPRET to the Study of Amazonian Jungle

The jungle of Peru is scarcely developed and its present conditions, in most of the amazon basin, is the same as it was millions of years ago¹⁵. The vegetation has evolved to adapt itself to the environment (soil, geomorphology, weather, etc.) and has reached the condition of climax forest in most of the Jungle. The rivers are the main dynamic agents and the wavelike evolution of the meanders create typical patterns with oxbow lakes and former river beds covered with herbaceous vegetation.

After discussions with professor Jorge Malleaux from the Forest Department of the Agrarian National University the area close to the town of Contamana on the Ucayali river was chosen, zone I of Fig.13. This region was attractive for this work because of the variety of forest types that presented.

At first, while INTERPRET was still in its preliminary form it was used for an identification of drainage patterns, lakes, man



1. The first step is to identify the variables involved in the problem. In this case, the variables are the number of hours worked (H) and the number of units produced (U).

ORIGINAL PAGE IS
OF POOR QUALITY

made clearings of the jungle or places where herbaceous vegetation has started to grow, swamps and wet areas.

In October 1974, under the auspices of the National Aerophotographic Service (SAN), a trip to zone I of Fig. 13 was made by personnel of ONERN and IGP. This mission flew over Contamana and over the region around the swamps south west of this town, black and white pictures were taken by an RC-10 camera and also 35 mm color Ektachrome pictures. This information and that already existent was used to help both the photointerpretation work at ONERN and the digital analysis at IGP.

Fig. 14 shows the result of an early version of INTERPRET made during the last part of 1974. Channels 5 and 6, channel 7 was missing due to unknown causes, on the scene of Fig. 3 were used. Fig. 14 is a computer output made with multiple print to obtain a gray level presentation. It was thought that this type of presentation would be advantageous for the delineation of the different classes.

Fig. 15 shows a color version of this analysis, it was obtained from computer printouts by coloring each character with different colors. This figure shows a branch of the Ucayali River west of Contamana and the swamp area south of it.

Blue, violet and black indicates different tones of water as seen from Landsat 1. The black represents highly sedimented water and the other colors, water with different concentrations of matter in suspension.



FIG.14. - Map from an early version of INTERPRET
Multiple print was used to obtain a
gray level presentation for the dif-
ferent classes.

The area covered are the swamps inside
zone I of Fig.13.

The green color represents a combination of water and vegetation, this was concluded after examining the ground truth and considering that the MSS detectors picks up an average of the reflected radiation over an area of approximately 5,000 m² and therefore a mixture of water and vegetation is observed at the border of lakes and rivers, it is interesting to observe that very narrow rivers like the one that goes from the swamps toward the north also appears green due to this effect on figure 15.

Orange represents zones where herbaceous vegetation is growing usually over former meanders or lakes. The dark yellow color represents swamp or muddy area.

Fig. 16 show a result from the present version of INTERPRET as it was displayed on the oscilloscope and Fig. 17 a display on the teleprinter. In the later picture, special characters designed to produce a gray level presentation were used.

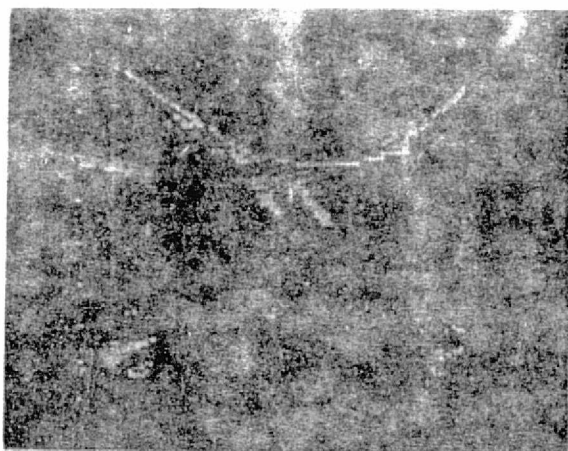
Fig. 18 shows a color version of the same result, here the jungle was also included, an attempt was made to match the classification obtained with INTERPRET to that obtained by Malleaux*from analysis of aerophotography. Much of the detail obtained with the computer was averaged out so that to have a more general presentation suitable for this comparison.

* Forestry Map of Peru 1976.

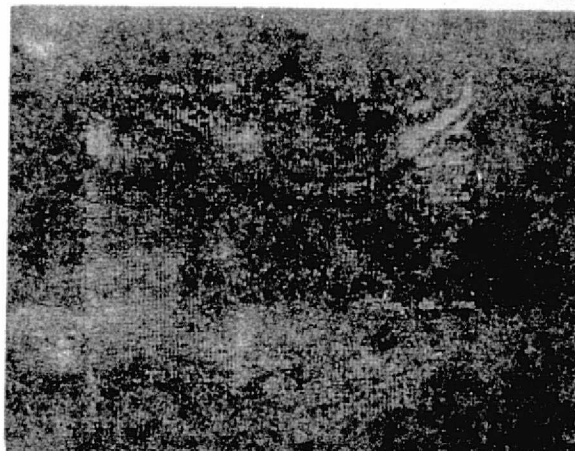
Fig.15 Color map from an early version of INTERPRET.

The following color code was used

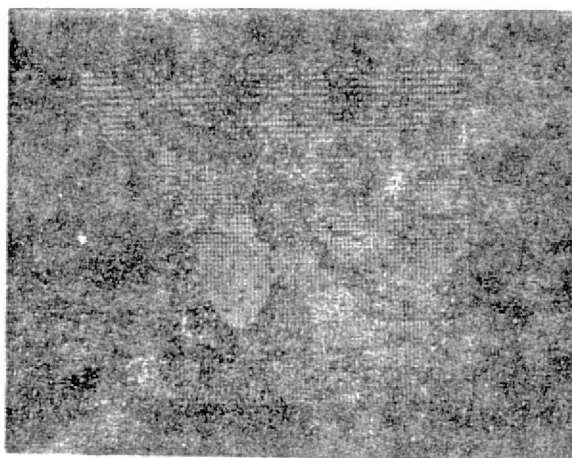
Black	highly sedimented water
Blue and Violet	water with intermediate degree of sedimentation
Orange.	herbaceous vegetation,
Green	mixture of water and vegetation, and wet areas
Dark yellow	swamps



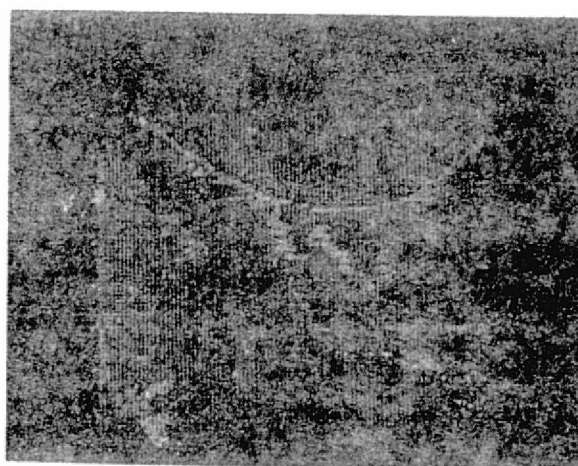
AGUA Y PANTANOS



AREA CON VEGETACION HERBACEA



AREA CON VEGETACION ARBOREA-
POBRE



COMPOSICION DE LAS 3 AREAS -
ANTERIORES

FIG. 16 - OSCILLOSCOPE MAPS FROM PRESENT
VERSION OF INTERPRET

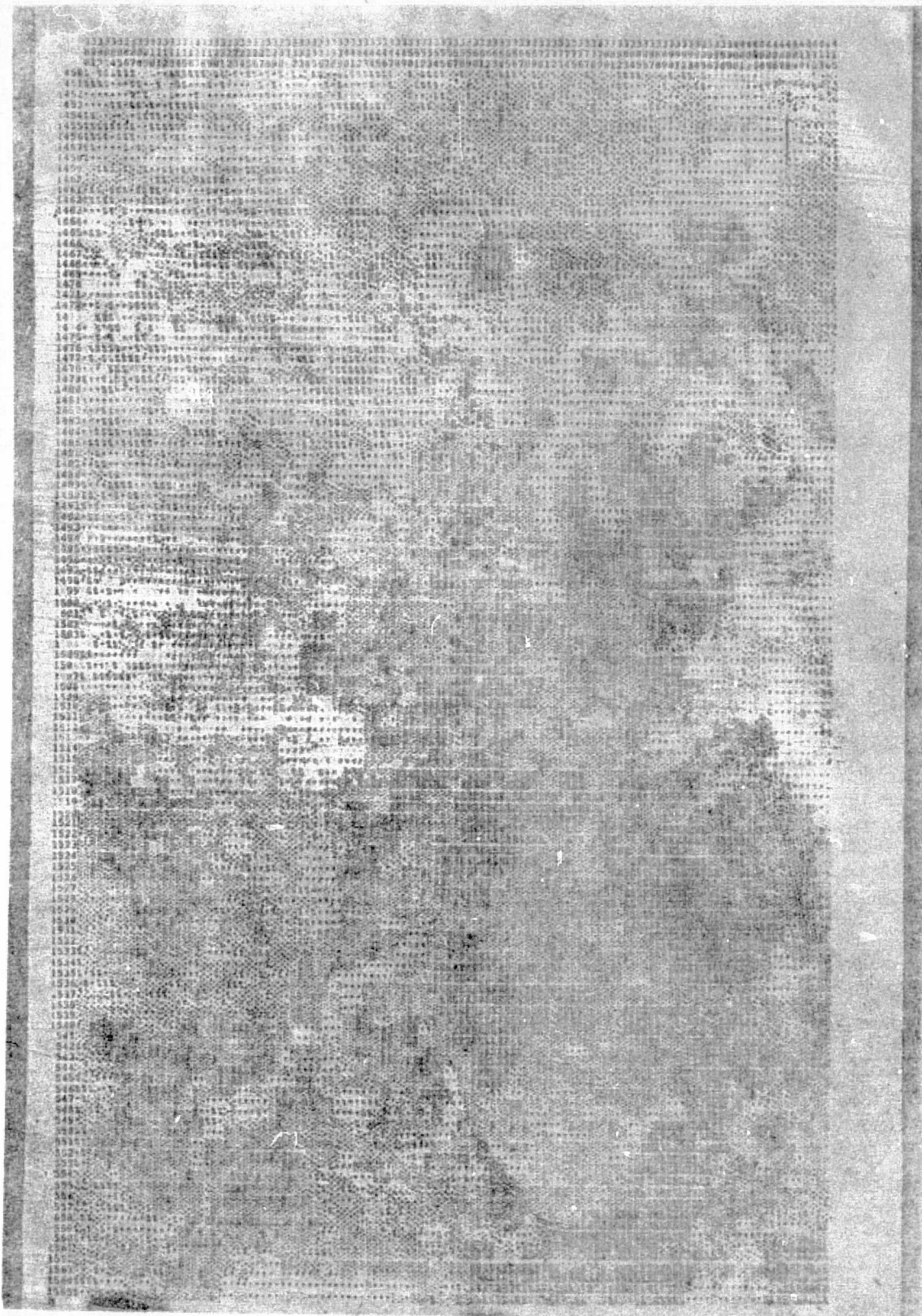


FIG.17 - Teleprinter map from the present version of INTERPRET.
Special characters are used for gray level presentation.

Fig. 18 Color map from present version of INTERPRET. The following color code was used:

Yellow	Alluvial forest type III
Green	Mixture of vegetation and water, wet areas or alluvial forest type II
Orange	Herbaceous vegetation or agricultural land
Blue	Water
Pink	Swamps
White	Unrecognized pixels

Fig. 19 shows an extract from the "Mapa Forestal del Perú" published by UNA at a 1:1,000,000 scale. A comparison with figure 18 shows that there are some obvious differences on the drainage patterns due to the fact that the aerophotography used as a base for the forestry map is over 20 years old while the Landsat imagery is recent, October 1972. It is interesting to observe that the lake east to the swamp area, zone I, has almost dried out and an oxbow lake has been formed south Contamana given a clear indication of the changes that are continuously occurring in the meander valley of the Ucayali river.

It can be seen that in spite of the averaging performed on the computer results to obtain figure 18, there is still a great deal of detail in this figure when compared with the forestry map.

The matching of figure 18 and 19 is satisfactory, the yellow color represents what has been defined as an alluvial forest type III on the forestry map.

The wet areas or mixture of vegetation and water are represented by green color in figure 18. It appears that the large area colored green at the left hand bottom corner of this figure may be also a representation of an Alluvial forest type II, this shall be investigated during the continuation of this research.

The orange color represents areas with herbaceous vegetation or also areas where the action of man has taken place. Blue color represents water and pink represents the swamp areas. White are those samples that were not recognized by the computer.

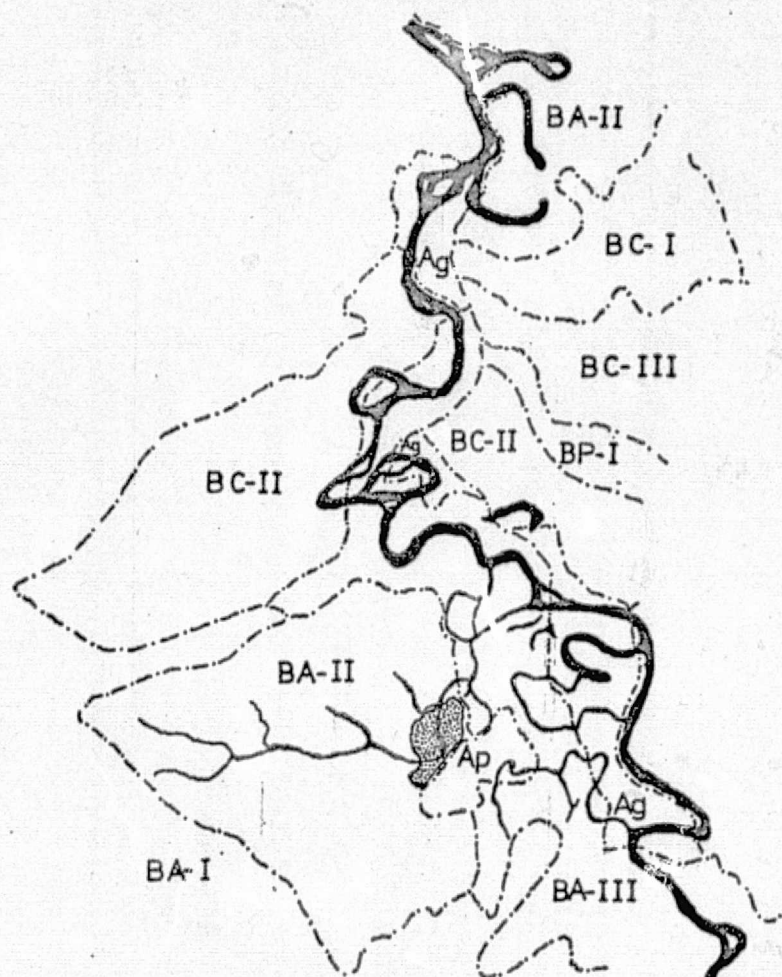


FIG. 19

- Ag AGRICULTURA
 Ap AREA PANTANOSA
 BA-I BOSQUE ALUVIAL I
 BA-II BOSQUE ALUVIAL II
 BA-III BOSQUE ALUVIAL III
 BC-I BOSQUE COLINA I
 BC-II BOSQUE COLINA II
 BC-III BOSQUE COLINA III
 BP-I BOSQUE PROTECCION I

EXTRACT FROM "MAPA FORESTAL DEL PERU 1976"
 SCALE 1:1,000,000
 NATIONAL AGRARIAN UNIVERSITY

When the areas south of Contamana are observed from an air plane they show an incipient development of agriculture which correlates very well with the orange area to the right of figure 18 South of the island and, extending toward the South. This is part of the area that is classified as agricultural land in the forestry map.

INTERPRET is also being tested on the study of the city of Tarapoto trying to determine the maximum amount of information that can be obtained from a Landsat imagery. Fig. 20 shows a preliminary result obtained on the oscilloscope from zone II. In Fig. 13 the city has been clearly delineated along with the main roads that converge on it.

At present there is not a recent map of the city and its surroundings, and it is probable that the present development of the agricultural areas close to Tarapoto have not been assessed yet. Classification maps from Landsat could be very valuable for the determination of the present state of development of cities like Tarapoto. This work with INTERPRET will continue in order to determine its actual potential as a tool for demographic studies.

VIII. Conclusions

The use of clustering methods permit the development of relatively fast classification algorithms that could be implemented in an inexpensive computer system with limited amount of memory.

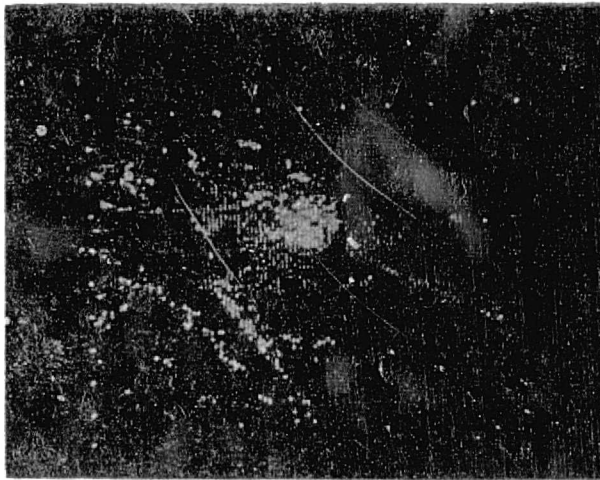


FIG. 20 - Map of Tarapoto. Brighter points represent urban areas. Medium level points represent unpaved roads. Weak level points represent water.

The analysis of CCTs by means of these techniques can provide a great deal of detail permitting the use of the maximum resolution of Landsat imagery.

Under the supervision of an specialized user the programs developed can produce satisfactory thematic maps in a relatively short time if an appropriate recording system is used.

In the course of this work it has been detected potential cases in which the use of other techniques for classification using a gaussian approximation for the distribution functions can be used with advantage. In general a system that uses several - different methods of classification is desirable.

For the study of jungle areas, channels 5 and 7 (or 6) can provide enough information to delineate drainage patterns, swamp and wet areas and make a reasonable broad classification of forest types.

The results obtained by means of INTERPRET were found satisfactory, by direct observation of the area from an airplane, aerophotography, and when compare with the Forestry Map published by UNA.

Computerized methods sustained by special purpose hardware can accelerate the process of evaluation of natural resources and represent the most logical way to deal with multiespectral imagery.

IX. Future Work

At present techniques based on the maximum likelihood ratio are being developed and will be added to INTERPRET.

Under the auspices of the Canadian International Development Agency the present computer system will be supplemented as shown in Fig. 21. This system will permit to produce thematic maps on film so that to permit enlargements to a scale appropriate for practical purposes.

Techniques for geometric correction, scale and earth rotation skew, shall be developed when the envisioned additions to the present computer system materialize.

Techniques based on linear and nonlinear adaptive algorithms are also being studied as additions to INTERPRET.

ORIGINAL PAGE IS
OF POOR QUALITY

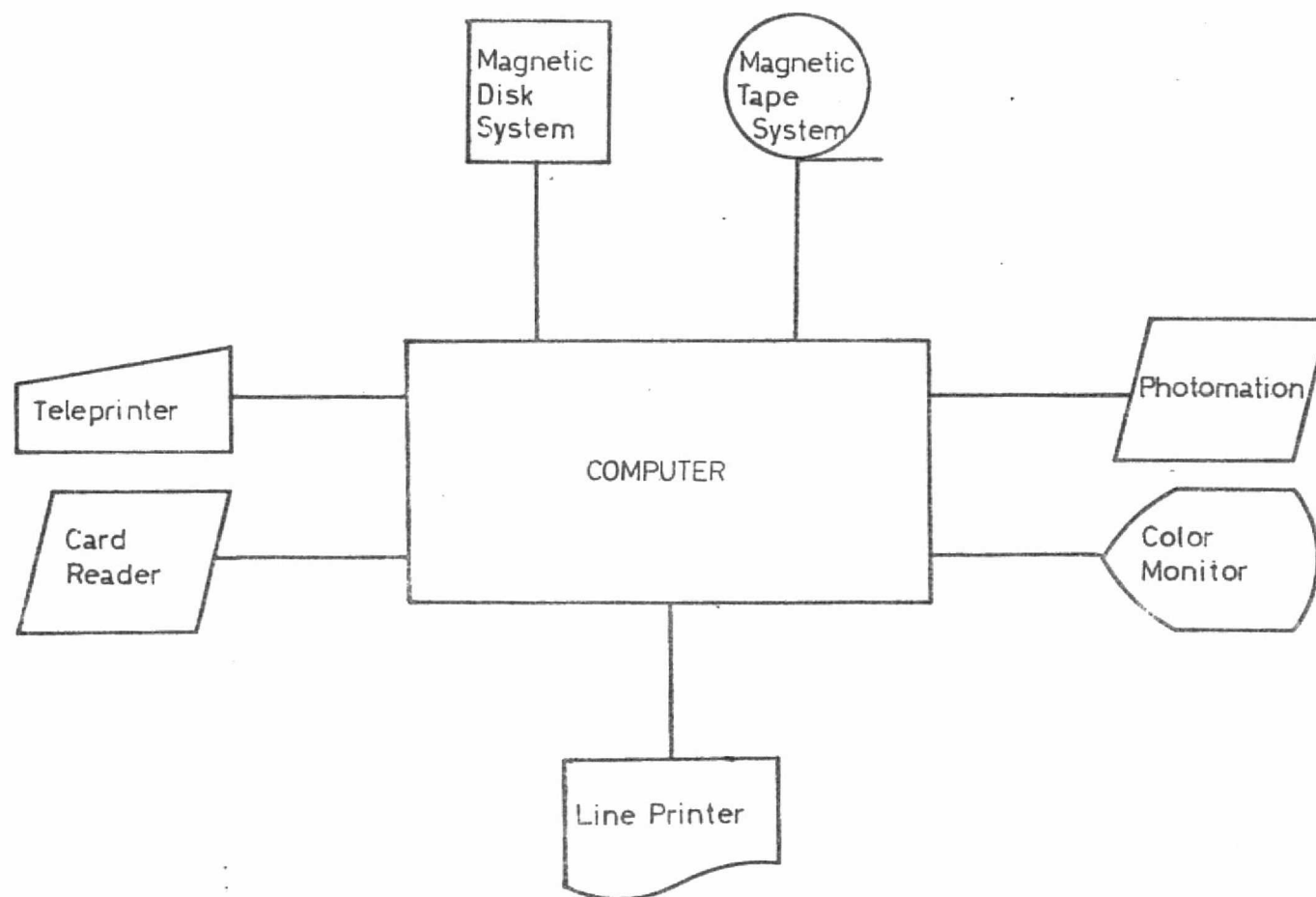


FIG.21- FUTURE COMPUTER SYSTEM FOR MACHINE ANALYSIS OF IMAGERY.

B I B L I O G R A P H Y

1. ERTS User's Manual, NASA
2. J. Malleaux "Estratificación Forestal con uso de Fotografías Aéreas" Vol. 1, Julio 1971. Universidad Agraria La Molina. Departamento Manejo Forestal.
3. Harry L. Van Trees "Detection, Estimation and Modulation Theory" Part. I, John Wiley, 1968.
4. Richard O. Duda, Peter Hart "Pattern Classification and Scene Analysis" John Wiley 1973.
5. Geoffrey H. Ball and David J. Hall "ISODATA, a Novel Method of Data Analysis and Pattern Classification" Stanford Research Institute, Menlo Park.
6. K. S. Fu "Information Processing of Remotely Sensed Agricultural Data" Proc IEEE, April 1969 - Vol.57 pp.639,652.
7. Y. C. Ho. "On Pattern Classification Algorithms Introduction and Survey" Proc IEEE, December 1968, pp.2101-2114
8. Paul E. Anuta, "Geometric Correction of ERTS-1 Digital Multispectral Scanner Data" Lars Purdue.
9. Geoffrey H. Ball "A comparison of Some Cluster Seeking Techniques" Stanford Research Institute, November 1966.
10. E. P. Kan "The JSC Clustering Program ISOCLS and ITS Applications" Lockheed Electronic Company, Inc. Houston Texas
11. T.M. Cover "Nearest Neighbor Pattern Classification" IEEE Transactions on Information Theory Vol. 13, January 1967, pp. 21-27
12. G. Sebestyen "An Algorithm for Non-Parametric Pattern Recognition" IEEE Transactions on Electronic Computers, Vol.15, December 1966, pp. 908-915.
13. F. Kriegler "Multivariate Interactive Digital Analysis System (MIDAS): A new Fast Multispectral Recognition System" Environmental Research Institute of Michigan (ERIM) Ann Arbor, Michigan.
14. Terry Phillips "The Implementation of the Maximum Likelihood Classification Rule, Assuming a Gaussian Density function" LARS Information Note, Purdue University.

15. Paul W. Richards "The Tropical Rain Forest", Scientific American, pp. 58-67, December 1973
16. K. S. Fu, D. A. Landgrebe, T. L. Phillips "Information Processing of Remotely Sensed Agricultural Data" Proc. IEEE Vol. 57, No.4, April 1969, pp. 639-653.
17. - W. G. Eppler Jr., R.P. Merrill Jr. "Relating Remote Sensor Signals to Ground-Truth Information Proc. IEEE, Vol. 57 No.4, April 1969.

Acknowledgement

This work has been carried out thanks to the help of various persons and organizations, and we are greatly indebted to all of them, specially to Mr. José Lizárraga, Director General of ONERN, and Mr. Alberto Giesecke, Director General of IGP, for their constant support and encouragement. Also, to the University of Engineering for the use of their Computer Center, to the National Service of Aerophotography for sponsoring our trip to the Contamana area, to professor Jorge Malleaux for his advice in the selection of the area of study, to Mr. Gilberto Tisnado for the modification of the Diablo Teleprinter for gray level presentation of our results and for his recommendations, and enthusiastic support. And finally to Stanford Research Institute for permitting the use of his digital equipment at Ancón.

2-46 a

A FORESTRY STUDY IN THE CONTAMANA - TARAPOTO
AREA BY MEANS OF LANDSAT I IMAGERY

Walter Danjoy

1976

Oficina Nacional de Evaluación
de Recursos Naturales

ABSTRACT

Landsat multiespectral images and aerophotography was used to classify the different jungle types in the region close to the cities of Contamana and Tarapoto. The attached map shows the results of this work.